



IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant: G. Paul Koning, Peter C. Hayden, Paula Long, and Daniel E. Suman
Application No.: 10/762,985 Group: 2157
Filed: January 21, 2004 Examiner: Sargon N. Nano
Confirmation No: 5999
For: Client Load Distribution

CERTIFICATE OF MAILING OR TRANSMISSION	
I hereby certify that this correspondence is being deposited with the United States Postal Service with sufficient postage as First Class Mail in an envelope addressed to Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450, or is being facsimile transmitted to the United States Patent and Trademark Office on:	
<u>9/23/08</u>	<u>Nadine Z. Kush</u>
Date	Signature
<u>Nadine Z. Kush</u>	
Typed or printed name of person signing certificate	

REPLY BRIEF

Mail Stop Appeal Brief-Patents
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

This Reply Brief is being filed in response to the Examiner's Answer mailed from the U.S. Patent and Trademark Office on July 23, 2008 in the above-identified application.

I. STATUS OF CLAIMS

Claims 1, 3-8, and 10-13 are pending.

Claims 1, 3-8, and 10-13 stand rejected and the rejections of these claims are being appealed.

A copy of the appealed claims appears in the Claims Appendix of the Appeal Brief filed May 13, 2008.

II. GROUND OF REJECTION TO BE REVIEWED ON APPEAL

(1) Claims 1, 3-8, and 10-13 were rejected under 35 U.S.C. § 102(e) by the Examiner as supposedly being anticipated by O'Neil et al., U.S. Patent Number 6,128,279 (hereinafter "O'Neil").

III. ARGUMENT

This Reply Brief addresses remarks presented in the Examiner's Answer dated July 23, 2008. Applicants discussion below addresses certain ones of the "Responses to Argument" in Section 10 of the Examiner's Answer.

(10) (A) O'Neil's "Load Balancing" Does Not Equate to Partitioning of Resources

The Examiner states that "O'Neil clearly divides the servers into two groups of servers logically and the resources of both groups are accessed by requests based on loads handled by these two groups of servers (see fig. 2A item # 203)." This is beside the point. Nowhere does O'Neil mention or suggest a plurality of storage servers having a set of resources partitioned thereon.

Figure 2A of O'Neil is referring to a load balancing function that distributes requests among network servers. (See O'Neil col. 6, lines 10-11.) For example, server 7 receives requests and then determines if the load currently being processed by it exceeds a first determined level. (See O'Neil col. 6, lines 10-11.) This level may be 50%, meaning that server 7 is operating at 50% capacity. (See O'Neil col. 6, lines 21-26.) At step S203, a decision is made whether the load currently being processed in server 7 exceeds a first predetermined level. (See O'Neil col. 6, lines 26-29.) The predetermined level is 50%, meaning that server 7 is operating at 50% capacity. *Id.*

If step S203 decides that server 7 is not processing a load that exceeds the first predetermined level, flow proceeds to step S204. In step S204, the network request is processed in server 7, and a response thereto is outputted via the appropriate channels. On the other hand, in a case that step S203 determines that server 7 is processing a load that exceeds the first predetermined level, flow proceeds to step S205. Step S205 determines loads currently being processed by server 7's peers (e.g., servers 9 and 10 shown in FIG. 3). In step S206, load balancing module 17 determines whether the loads currently being processed by server 7's peers are less than the load on server 7 by a differential exceeding a second predetermined level. The second predetermined level is 20%, which provides a means of assessing whether servers 9 or 10

have at least 20% more of their capacities available than server 7. (*See* O’Neil col. 6, lines 30-55.)

Thus, O’Neil admittedly is balancing load among different servers. However, this does not equate to partitioning resources as claimed by Appellants. In order to determine what is meant by the Applicants’ claim “plurality of storage servers having the set of resources partitioned thereon”, the Patent Office is reminded that one must look to the Applicants’ own specification to determine the meaning of a claim. While we do understand that limitations from the specification should not be read into the claims, when the specification specifically states the meaning that Applicant intended particular words to have, the claim is examined using that meaning, in order to achieve a complete exploration of the Applicants’ invention in its relation to the prior arts. *In re Zletz*, 893 F.2d 319, 321-22 (Fed. Cir. 1989).

Applicants’ partitioned-resource storage system can be more particularly understood from the illustrative embodiment of Figs. 1 and 3. Fig. 1 depicts client 12 communicating with a plurality of servers 161, 162 and 163 (collectively, server group 16) to obtain access to a resource. In this example, the resource partitioned across the server group 16 may be a data block, page, file, database, application, or other resource. For illustration, the diagram here shows two resources, one resource 18 that is partitioned over all three servers, servers 161, 162, 163. Another resource 17 is partitioned only over two of the three servers. Partitioning of a resource requires dividing it up in some way so that certain parts are available in one place only and other parts in another place.

Referring to Fig. 3, each server in the group 16 thus needs a routing table 165 for each volume. The routing table 165 identifies the particular server on which a specific resource (e.g., a page of a specific volume) can be found. For example, when the server 161 receives a request from a client 12 for a specific resource (volume 3, block 93847), the server 161 calculates the page number (page 11 in this example for the page size of 8192) and looks up in the routing table 165 the location or number of the server that contains page 11 (in the example discussed, this is server 163). It is important to note that only server 163 has the requested resource because of the partitioning.

Therefore, O’Neil is simply load balancing among multiple identical servers, each of which can provide the complete resource requested. On the other hand, Applicants

claimed “partitioning of the resources” requires that some parts of a resources are available in one place (e.g., only on one server) and another portion of that resource is available only in another place (e.g., another server).

Applicants therefore require a structure such as, for example, a routing table 165 to identify where a each particular portion of a resource can be located. It isn’t enough, for example, to try to locate the requested resource in Applicants’ system such as by locating the “least busy server” as does O’Neil. Such would not result in locating the requested resource. Rather, because Applicants partition resources, the resource request must ultimately be sent to a server that has that particular portion of the resource that has been requested.

There is simply no way that load-balancing alone as taught by O’Neil is the same thing as Applicants’ claimed “plurality of storage servers having the set of resources partitioned thereon.”

The Examiner’s statement that “each server has a partitioned memory” in O’Neilm, with reference to column 6, lines 50-61, makes no sense at all. All that section states is that in step S206 a load-balancing module 17 determines whether the load currently being processed on a given server are less than the load on other servers by certain predetermined amount(s). O’Neil refers to memory locations here only in the context of a place to store the value for the predetermined amount(s), and that the value can be reprogrammed. This also does not equate to partitioning resources at all.

(10) (B) Redirecting a Request to a Less Loaded Server is Not Identical to Moving an Already Established Connection

The Examiner also argues that that “O’Neil discloses a load balancing technique where a client request is redirected or moved to a less loaded server.” Applicants respectfully disagree.

Briefly, Applicants’ client-server connections are long-lived and typically last across periods of time during which different system and client loads are experienced by the group and respective servers. As a result, any initial client-server assignment for a new connection based on load (see, e.g., Specification at page 18, line 18 to page 19, line 8) may subsequently not yield

the best overall or server-specific performance. To address this issue, the Applicants' system periodically "redistributes" existing client connections based on load. For example, in Fig. 5, a client distribution process 30 moves the connection for client 3 (12C) connection from server 161 to server 162 based on load measured by a distributed load monitoring process. (See, e.g., Specification at page 20, lines 14-17.) Once the connection is moved, server 162 services requests for client 3 (12C) requests (which may involve accessing other servers for portions of the resource that are not stored on server 162).

In contrast, O'Neil merely teaches that the client requests are re-routed to particular servers. (See, O'Neil col. 4, lines 4-13.) O'Neil makes it clear that this re-routing of requests does not involve moving a given connection from one server to another, but rather involves "breaking" the connection and sending a message that redirects the client to the new server. For example, "In the invention, routing is performed by sending a command from load balancing module 17 to a requestor instructing the requestor" (e.g., by sending an instruction to the client) "to send the request to a designated server." (See O'Neil column 7, lines 26-30; column 8, lines 32-33; column 8, lines 51-53; and column 9, lines 2-4.) This is not moving of already established connections in O'Neil as required in Applicants' claimed invention. In a partitioned storage system, connections may last much longer and across disparate loading conditions. Therefore, the capability of moving connections from one server to another must be accommodated. Furthermore, we note that with Applicants' claimed repartitioning and moving of the set of connections, the clients are not involved in the process at all. Therefore, how can they be the same?

(10) (D) A Network Server is Not a Storage Area Network

The Examiner states that "O'Neil discloses a plurality of network servers which directly monitor and handle load balancing on request received." Applicants respectfully disagree. The Examiner fails to show where O'Neil teaches each and every element of Applicants' Claim 13 identically as claimed.

Nowhere does O'Neil teach, mention, or even suggest the use of his servers to provide a Storage Area Network (SAN). O'Neil's stated purpose is instead to provide a web server.

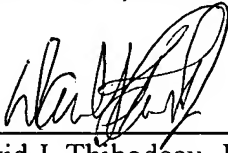
For example, in the embodiment shown in Fig. 5 and explained at page 15, lines 5-14 of Applicants' specification, the server group 16 may be a Storage Area Network (SAN) or part of a SAN, wherein each of the equivalent servers has an individual IP address that maybe employed then by a client for accessing that particular equivalent server. To that end the routing tables of the individual equivalent servers coordinate with each other to provide a global database of the different storage resources that they provide. In the example embodiment of Fig. 5, those resources include data blocks, pages or other organizations of data blocks. Web servers do not provide such storage area network functionality.

CONCLUSION

In view of the foregoing and the arguments previously presented in the Appeal Brief filed on May 13, 2008, reversal of the rejections is respectfully requested.

Respectfully submitted,

HAMILTON, BROOK, SMITH & REYNOLDS, P.C.

By 

David J. Thibodeau, Jr.
Registration No. 31,671
Telephone: (978) 341-0036
Facsimile: (978) 341-0136

Concord, MA 01742-9133

Date: 9/23/08